

Anpassung von individuellen Mediennutzungswahrscheinlichkeiten

Kopp, Bruno

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Kopp, B. (2011). Anpassung von individuellen Mediennutzungswahrscheinlichkeiten. *transfer Werbeforschung & Praxis*, 57(2), 6-15. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-254164>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:
<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:
<https://creativecommons.org/licenses/by/4.0>

Anpassung von individuellen Mediennutzungswahrscheinlichkeiten

**Bruno Kopp**

Selbstständiger mathematischer Berater, Schneverdingen

✉ bruno-kopp@t-online.de

Es wird eine Anpassungsmethode von Nutzungswahrscheinlichkeiten bei Umfragen zur Mediennutzung untersucht. Das Verfahren ist so angelegt, dass die Reichweiten in vorgegebenen Merkmalen (Zielgruppen) an Sollwerten justiert werden. Ein Kontrollparameter sorgt bei dieser Untersuchung dafür, dass die ursprünglichen Nutzungswahrscheinlichkeiten dabei möglichst wenig verändert werden. Damit der vorgestellte Algorithmus auch allgemeinverständlich bleibt, wird hier auf Formeln weitgehend verzichtet und die Vorgehensweise an einem Rechenbeispiel erläutert.

Eingereicht: 02.03.2010, akzeptiert: 14.07.2010

Schlagworte: > Anpassung > Mediennutzung > Mediennutzungswahrscheinlichkeiten > Mediaanalyse

1. Einführung

Sämtliche Medien (Zeitungen und Zeitschriften für den Printbereich und TV / Rundfunk bei den elektronischen Medien) in der Bundesrepublik Deutschland werden in der jährlich wiederkehrenden Befragung der Media-Analyse (MA) der Arbeitsgemeinschaft Media-Analyse e.V. (ag.ma) erfasst. Die Ergebnisse der MA gelten als gültige Norm für die Werbung treibende Wirtschaft. Dabei wurden zum Beispiel im Jahr 2010 (für den Printbereich) insgesamt 38.903 Personen repräsentativ für die Grundgesamtheit der Bevölkerung der BRD ab 14 Jahren in Privathaushalten (64,82 Mio.) befragt. Ein wesentlicher Bestandteil der Untersuchungen sind die individuellen Nutzungswahrscheinlichkeiten eines jeden Befragten zu den erfassten Medien. Diese werden rechnerisch für jedes Medium per Segmentierung aus den erhobenen Daten ermittelt. Alle Befragten eines Segments erhalten denselben „p-Wert“ als Quotient der Summe aller Nutzer im letzten Erscheinungsintervall (Print Medien) zur Gesamtanzahl der Befragten im Segment. Da sich nach jedem Schritt die Anzahl der Segmente verdoppelt, erhält man bereits nach vier Segmentationsstufen 16 unterschiedliche Gruppen. Üblicherweise wird nach Geschlechtern getrennt segmentiert; dadurch erhält man dann (maximal) 32 unterschiedliche p-Werte. Diese p-Werte sind Bausteine für wichtige Kenngrößen der Medien, wie deren Reichweite als Summe der (gewichteten) p-Werte. Die MA

liefert also im Ergebnis die Reichweiten als Kenngröße der abgefragten Medien. Daraus resultiert die Notwendigkeit der p-Wert-Anpassung bei anderen Befragungen mit gleichem Medienhintergrund, weil die Werbung treibende Wirtschaft auf einen einheitlichen Standard bei Maßzahlen wie die Reichweite in besonderer Weise angewiesen ist.

So machen auch Verlage eigene repräsentative Befragungen derselben Grundgesamtheit bei freilich wesentlich kleineren Stichproben. Allerdings bricht bei solchen Untersuchungen das Segmentationsverfahren bereits nach wenigen Segmentationsstufen ab, woraus sich eine zu grobe Struktur der p-Werte ergibt. Dies gilt auch dann, wenn man die Segmente durch andere Gruppierungsarten ersetzt, etwa über Frequenzklassen (Anzahl genutzter Ausgaben eines Mediums), getrennt nach Geschlecht bzw. Alter. Daraus folgt zwangsläufig die Notwendigkeit einer nachträglichen Korrektur zur Anpassung der p-Werte an die gültigen Normen, also die Reichweiten der MA als Sollwerte. Gemeint ist dabei, dass die Reichweiten in elementaren Zielgruppen (Geschlecht bzw. Alter) zur Übereinstimmung gebracht werden, wobei die ursprünglichen Nutzungswahrscheinlichkeiten nur marginal verändert werden dürfen. Da in dieser Arbeit ausnahmslos mit fiktiven Daten gearbeitet wird und um Missverständnissen vorzubeugen, ist hier statt von der MA-Untersuchung von der Referenzuntersuchung REF die Rede. Wenn also von Sollwerten gesprochen wird, so sind damit grundsätzlich die aus dieser REF-Untersuchung abgeleiteten Werte gemeint.



Wir wissen, was Sie letzte Woche gesehen haben!

ERREICHEN SIE MIT IHREM KINOSPOT IHRE KUNDEN JETZT NOCH GENAUER!

Denn mit **cinecom Ticketing** gibt es jetzt ein brandneues, international einzigartiges Tool, mit dem sich die Kinobesucher österreichweit genau analysieren lassen. Die Auswertung der Daten ermöglicht präzise Zielgruppenprognosen für alle neu anlaufenden Filme - und gibt Ihnen noch mehr Planungssicherheit für Ihre Buchungen auf dem **cinechannel**.

Kleine Ursache, GROSSE WERBEWIRKUNG.

Überzeugen Sie sich von **cinecom Ticketing** und den noch zielgruppenaffineren Buchungen über den **cinechannel**.

Infos erhalten Sie unter **01/587-05-15-0**
oder unter **www.cinecom.at**

Es werden nun nacheinander drei Anpassungs-Methoden entwickelt. Die erste Methode hat gegenüber der zweiten Methode den Nachteil, dass sie unter noch zu präzisierenden Umständen versagen kann. Während die Ergebnisse der Methoden 1 und 2 sich als geschlossene Form einer Gleichungslösung darstellen lassen, arbeitet Methode 3 iterativ und vermeidet alle Mängel der vorgenannten Verfahren.

Um die Lektüre hier nicht über Gebühr zu strapazieren, werden lediglich im Anhang die wichtigen mathematischen Formeln aufgeführt. Dieser Anhang ist für den eher technisch interessierten Leser gedacht. Die dort angegebenen Beziehungen sind eine Weiterentwicklung der Formeln eines früheren Aufsatzes (Kopp 2006). Im Übrigen gibt der Beitrag auch ohne tiefe mathematische Kenntnisse interessante Hinweise zu der Problematik der p-Wert-Anpassung, für die es in der Literatur kaum Fundstellen über die Anpassungstechnik gibt. Es soll hier noch kurz auf ein heuristisches Verfahren eingegangen werden, dessen Zielsetzung in erster Linie auf die Verteilung der p-Werte gerichtet ist (Büttner 1979). Bei diesem Verfahren werden die p-Werte einer MA an diejenigen einer anderen (früheren) MA angepasst, damit die Reichweiten in der Gesamtbevölkerung übereinstimmen. Dies geschieht in einem zweistufigen Prozess. Die erste Stufe sorgt dafür, dass die Verteilungen der p-Werte in den sechs Klassen bei beiden Befragungen übereinstimmen.

Nichtnutzer des Mediums:	0,00 – 0,00
ganz seltene Nutzer:	0,01 – 0,24
seltene Nutzer:	0,25 – 0,41
gelegentliche Nutzer:	0,42 – 0,58
häufige Nutzer:	0,59 – 0,82
Kernnutzer:	0,83 – 1,00

Dieser Prozess lässt sich modelltechnisch als Lösung eines sogenannten modifizierten Transportproblems interpretieren, bei dem die p-Werte der aktuellen MA so lange umgruppiert werden, bis die Gleichheit der Verteilung erreicht ist (Müller-Merbach 1973).

Das führt in einem ersten Schritt dazu, dass sich für die anzupassenden p-Werte zunächst nur Bereiche (s. o.) aber noch keine konkreten Werte ergeben (Stufe a). Erst in Stufe b erfolgt daraus die endgültige Bestimmung der p-Werte über ein heuristisches Verfahren innerhalb der gefundenen Bereiche. Verteilungsinvarianz der p-Werte bei unterschiedlichen MA-Untersuchungen mag zwar sinnvoll und wegen der großen Stichproben auch einfach zu realisieren sein. Sie ist jedoch bei kleineren Untersuchungen zumindest problematisch, weil hier Gruppenwechsel häufiger auftreten würden und damit der Vorwurf unzulässiger Manipulation verbunden sein könnte. Ein Wechsel vom Kernnutzer zum häufigen Nutzer könnte zum Beispiel zur Folge haben, dass ein ursprünglicher p-Wert von 1,00 auf 0,59 herabgestuft

wird. Nicht auszuschließen sind auch extreme Wechsel, etwa vom Kernnutzer zum ganz seltenen Nutzer, mit einer p-Wert-Reduktion von 0,99.

Im Unterschied dazu liefern die nachfolgenden Methoden Übereinstimmungen nicht nur in der Gesamtbevölkerung, sondern auch in beliebigen Zielgruppen. Außerdem handelt es sich hierbei um mathematische Optimierungsverfahren, die auf Heuristiken gänzlich verzichten.

2. Problemstellung - Datenmaterial zur Demonstration: Eine fiktive Befragung von 20 Personen

Die Reichweite eines Mediums ist definiert durch $\sum_{i=1}^n g_i p_i$

Dabei sind g_i die Gewichte der Befragten und p_i deren Nutzungswahrscheinlichkeiten zwischen 0 und 1. Summiert wird über die Anzahl n der Befragten. In unserem Beispiel ist $n = 20$ und sämtliche Gewichte sind gleich 1. Die im Anhang aufgestellten Formeln sind natürlich allgemeingültig für beliebige Gewichte. Summiert man in obiger Formel nur über Befragte einer Zielgruppe (z. B. Männer einer bestimmten Altersgruppe), so liefert die Formel die Reichweite in dieser Zielgruppe. Bei der obigen Reichweite handelt es sich um eine absolute Zahl, die sich hier ausdrücklich auf die Größe der Befragung ($n = 20$) bezieht. Die tatsächliche Reichweite bezogen auf die Bevölkerung erhält man durch Multiplikation mit dem Hochrechnungsfaktor Bevölkerungsgröße / n . Darauf wird hier verzichtet, weil dies für die Vorstellung der Anpassungsverfahren ohne Belang ist.

In nachstehender **► Tabelle 1** sind in den Spalten 2 bis 7 die Werte für 20 „Befragte“ einer fiktiven Untersuchung gelistet. Hierbei handelt es sich um das Kombinationsmerkmal Geschlecht x Alter.

Ursprünglich hatte das Alter drei Ausprägungen. Die Variable Geschlecht x Alter wurde in Alternativmerkmale aufgelöst: m, 1 bzw. m, 3 bedeuten Männer in den Altersstufen 1 bzw. 3. Entsprechendes gilt für die Binärmerkmale bei den Frauen, so dass insgesamt sechs Zielgruppen ZG 1, ZG 2, ZG 3, ZG 4, ZG 5 und ZG 6 entstehen. Am Fußende unterhalb der sechs Binärmerkmale sind die Sollwerte der Reichweiten in diesen Zielgruppen angegeben. Erläuterung: Diese Sollwerte ergeben sich aus der Referenzuntersuchung REF: Ist N die Stichprobengröße von REF, R der Wert einer Reichweite (gemessen als Fallzahl der N Befragten von REF), so ergäbe sich der entsprechende Sollwert für die 20 Befragten per Kalibrierung zu $R \times (20/N)$. So bedeutet 4,32 die geforderte Reichweite eines Mediums bei den 20 Befragten für die Zielgruppe der Männer in Altersstufe 1, entsprechend gilt für die Reichweite bei den Frauen, Altersstufe

Tab. 1: Fiktive Befragung von 20 Personen

(1) Befragter	Zielgruppe						Ursprüngl. Nutzungswahr- scheinlichkeit		Durch die drei Modelle berechnete Nutzungswahr- scheinlichkeit	
	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	ZG 1	ZG 2	ZG 3	ZG 4	ZG 5	ZG 6	p0	p1	p2	p3
	m, alt 1	m, alt 2	m, alt 3	w, alt 1	w, alt 2	w, alt 3				
1	1	0	0	0	0	0	0,65	0,92	0,91	0,96
2	1	0	0	0	0	0	0,37	0,64	0,63	0,68
3	0	0	1	0	0	0	0,79	0,55	0,57	0,57
4	0	0	1	0	0	0	0,84	0,6	0,62	0,62
5	0	0	0	0	1	0	0,19	0,42	0,39	0,44
6	0	0	0	0	0	1	0,54	0,69	0,68	0,68
7	1	0	0	0	0	0	0,40	0,67	0,66	0,71
8	1	0	0	0	0	0	0,58	0,85	0,84	0,89
9	1	0	0	0	0	0	0,97	1,24	1,23	1,00
10	0	1	0	0	0	0	0,53	0,43	0,43	0,43
11	0	1	0	0	0	0	0,95	0,85	0,85	0,85
12	0	0	0	0	1	0	0,86	1,09	1,06	1,00
13	0	0	0	0	0	1	0,44	0,59	0,58	0,58
14	0	0	0	0	0	1	0,71	0,86	0,85	0,85
15	0	1	0	0	0	0	0,83	0,73	0,73	0,73
16	0	0	0	0	0	1	0,77	0,92	0,91	0,91
17	0	1	0	0	0	0	0,61	0,51	0,51	0,51
18	0	0	0	1	0	0	0,93	1,15	1,13	1,00
19	0	0	1	0	0	0	0,83	0,59	0,61	0,61
20	0	0	0	1	0	0	0,54	0,76	0,74	0,84
	Soll	Soll	Soll	Soll	Soll	Soll		ABW*	ABW*	ABW*
	4,32	2,51	1,74	1,92	1,5	3,06		0,87	0,76	0,83

* ABW: Abweichung

3, 3,06. Alle Angaben der Reichweiten beziehen sich hier und im Folgenden auf die Anzahl $n = 20$ der Befragten. Diese Sollwerte stammen – wie erwähnt – aus der entsprechenden Referenzuntersuchung. Beispiel: Im Falle REF = MA2010 würden die dortigen absoluten Reichweiten in den sechs Zielgruppen (mit dem Kalibrierungsfaktor $20 / 38.903$ multipliziert) diese Sollwerte für die fiktive Befragung liefern.

Die Addition dieser sechs Sollwerte ergibt den Wert 15,05. Dies ist gleichzeitig die Gesamtreichweite (als Sollwert), weil die sechs Zielgruppen überschneidungsfrei und exhaustiv (also vollständig) die 20 Befragten überdecken.

Unter der Überschrift „p0“ in Spalte 8 befinden sich die ursprünglichen p-Werte der Befragung, also die Nutzungswahrscheinlichkeiten vor der Anpassung. Die mit diesen Originalwerten erzielten Reichweiten innerhalb der Zielgruppen lassen sich aus [Tabelle 1](#) einfach ermitteln. So gilt für ZG 1: Die Befragten 1, 2, 7, 8 und 9 in dieser Gruppe haben die p-Werte (vgl. Spalte 8) 0,65, 0,37, 0,40, 0,58 und 0,97. Daraus ergibt sich als Reichweite für ZG 1 auf Basis der ursprünglichen p-Werte: $0,65 + 0,37 + 0,40 + 0,58 + 0,97 = 2,97$. Analog erhält man für ZG 2: 2,92, ZG 3: 2,46, ZG 4: 1,47, ZG 5: 1,05 und für ZG 6: 2,46. Demzufolge liefern die Originalwerte eine Gesamt-

reichweite von 13,33. All diese Werte sind die Reichweiten, wie sie noch vor der Anpassung (also aufgrund der Originalwerte in p_0) anfallen. Ziel ist es, diese Werte auf die Sollwerte zu transformieren. Insbesondere soll die Gesamtreichweite von 13,33 auf 15,05 gebracht werden. Die Überschriften „p1“, „p2“ und „p3“ beziehen sich auf angepasste p-Werte als Lösungen der eingangs erwähnten Verfahren, die nun der Reihe nach vorgestellt werden sollen.

3. Drei Lösungsansätze

3.1 Die Methode der parameterfreien, direkten Anpassung der p-Werte („p1“)

Bei dem ersten Ansatz wird ein nicht-iteratives Verfahren (d. h. das Ergebnis wird durch eine Formel rechnerisch zur Verfügung gestellt) benutzt. Es hat, um es gleich vorweg zu nehmen, zwei Nachteile. Zunächst können p-Werte auftreten, die außerhalb des Bereichs von 0 bis 1 liegen. Außerdem,

was noch gravierender ist, kann es, in Abhängigkeit von der Datenlage vorkommen, dass die benutzte Formel (im Anhang ist dies Nr. 2) versagt. Dies ist genau dann der Fall, wenn nicht nur ein einziges Merkmal wie in unserem Beispiel (Geschlecht x Alter) verwendet wird. Würde man noch ein weiteres hinzufügen, etwa das Nettoeinkommen mit all seinen Ausprägungen (stufenweise als Alternativmerkmale ausgewiesen: Stufe 1: Ja = 1, Nein = 0, etc...), so ließe sich damit die Formel 2 nicht mehr anwenden. Dieser Fall ist in den Demonstrationsdaten (nur ein Merkmal) hier ausgeschlossen.

Methode 1 bestimmt eine Lösung p1 derart, dass diese Reichweiten in den sechs Zielgruppen ZG 1 (Männer, Altersstufe 1) bis ZG 6 (Frauen, Altersstufe 3) generiert, die mit ihren Vorgabewerten (Soll) grundsätzlich identisch sind. Diese Forderung - Übereinstimmung der Reichweiten in den sechs Zielgruppen - hat mathematisch unendlich viele Lösungen. Von all diesen theoretisch denkbaren wird nun genau diejenige Lösung p1 ausgewählt, die möglichst

Tab. 2: Soll-Ist-Vergleiche der Reichweiten für drei Methoden

Methode	Zielgruppe	Soll	Ist	Abweichung der p-Werte von p_0
(1)	(2)	(3)	(4)	(5)
1	ZG 1	4,32	4,32	
	ZG 2	2,51	2,52	
	ZG 3	1,74	1,74	
	ZG 4	1,92	1,91	
	ZG 5	1,50	1,51	
	ZG 6	3,06	3,06	
	Gesamt	15,05	15,06	0,87
2	ZG 1	4,32	4,27	
	ZG 2	2,51	2,52	
	ZG 3	1,74	1,80	
	ZG 4	1,92	1,87	
	ZG 5	1,50	1,45	
	ZG 6	3,06	3,02	
	Gesamt	15,05	14,93	0,76
3	ZG 1	4,32	4,24	
	ZG 2	2,51	2,52	
	ZG 3	1,74	1,80	
	ZG 4	1,92	1,84	
	ZG 5	1,50	1,44	
	ZG 6	3,06	3,02	
	Gesamt	15,05	14,86	0,83

Akku-Geräte aus
Österreich!



Im siebenten Himmel!

Gemeinsame Kraftquelle für STIHL und VIKING.

Mit der Akkutechnologie von STIHL werden neue Maßstäbe gesetzt. Die leistungsstarken, geräuschlosen und geruchsfreien Akkugeräte von STIHL sind vielseitig einsetzbar. Besonders für lärmsensible Bereiche wie Innenhöfe,

Dachterrassen, Gärten oder auch in Innenräumen sind sie perfekt geeignet. Das System ist einfach – ein Akku für alle Geräte – alle Geräte mit einem Akku. Ob Motorsäge, Heckenschere, Motorsensen oder Rasenmäher – für jede Anwendung das richtige Gerät. Die einfache Handhabung,

die hohe Leistungsstärke, die exakte Ladeanzeige und die volle Akku-Leistung bis zur nächsten Aufladung sorgen für eine perfekte Zeiteinteilung und Spaß an der Arbeit.

STIHL Akku-Geräte werden in Österreich hergestellt. Erhältlich im STIHL Fachhandel.



Standardladedegerät AL 100

Für STIHL Akkus
AP 80, AP 120, AP 160



Schnellladedegerät AL 300

Für STIHL Akkus
AP 80, AP 120, AP 160

wenig von den mit p_0 bezeichneten Originalwahrscheinlichkeiten abweicht. Durch diese Zusatzforderung wird die Lösung p_1 eindeutig. Das Verfahren ist ein quadratisches Optimierungsproblem mit Gleichungen als Nebenbedingungen. Der obere Teil von **► Tabelle 2** bestätigt folgerichtig die Übereinstimmung (bis auf Rundungsfehler) zwischen den aus REF abgeleiteten Reichweiten (Soll) und den von Methode 1 generierten Reichweiten (Ist).

Zur Erläuterung soll der Ist-Wert 4,32 in **► Tabelle 2**, also die vom Verfahren generierte Reichweite bei ZG 1 rechnerisch nachvollzogen werden. Man beachte, dass in **► Tabelle 1** den fünf Einsen in ZG 1 die p_1 -Werte 0,92, 0,64, 0,67, 0,85 und 1,24 gegenüberstehen. Addiert ergeben diese die Reichweite von 4,32. 4,32 ist also die nach durchgeführter Anpassungsmethode 1 erhaltenen Reichweite. Diese muss – bis auf Rundungsfehler – nach obigen Ausführungen mit ihrem Soll-Wert übereinstimmen! Dies gilt genauso für die angepassten Reichweiten ZG 2 bis ZG 6 und ist das Charakteristikum von Methode 1: Soll-Spalte = Ist-Spalte im oberen Teil von **► Tabelle 2**. Das Maß für die Abweichung der von Methode 1 angepassten p -Werte p_1 von den Originalwerten in p_0 beträgt 0,87 und ist in **► Tabelle 1** unten und in **► Tabelle 2** in Spalte 5 eingetragen. Später wird sich zeigen, dass die Methoden 2 und 3 hier bessere (also kleinere) Abweichungswerte liefern. Ein Nachteil von Methode 1 wird in **► Tabelle 1** bei p_1 ersichtlich: Die Befragten 9, 12 und 18 haben unzulässige p -Werte über 1 (1,24, 1,09 und 1,15).

3.2 Die Methode mit Kontrollparameter und direkter Anpassung der p -Werte

Dieses Verfahren ist eine Weiterentwicklung von Methode 1. Im Gegensatz zum vorhergehenden Verfahren funktioniert diese Methode unabhängig von der Anzahl eingesetzter Merkmale, d. h. hier würden auch die Variablen Geschlecht \times Alter und zusätzlich das Nettoeinkommen keine Probleme machen. Nach wie vor aber können unzulässige Wahrscheinlichkeitswerte nicht ausgeschlossen werden. Bei dieser Methode wird auf die exakte Übereinstimmung der Reichweiten bei Soll und Ist verzichtet. Vielmehr gewährleistet ein positiver Kontrollparameter r eine flexible Abweichungstoleranz zwischen Soll und Ist. Wenn Z_1 die Summe der quadrierten Abweichungen zwischen den Soll-Reichweiten und den angepassten Reichweiten bezeichnet, und Z_2 für die Summe der quadrierten Abweichungen zwischen den angepassten p -Werten und ihren Originalwerten in p_0 steht, so wird bei dieser Methode die Zielfunktion $Z = r Z_1 + Z_2$ minimiert. Hieran erkennt man, dass Z_1 mit dem Gewicht r in die Optimierung eingeht. Wachsendes r begünstigt die Anpassung der Reichweiten (die Abweichung Z_1 konvergiert gegen 0). Sinkende Werte von r führen dazu, dass die Lösung p_2 näher an den Originalwerten p_0 liegt (die Abweichung Z_2 geht gegen 0). Die im Anhang angegebene Formel 5 ist die Lösung einer (unbeschränkten) Optimierungsaufgabe.

Im mittleren Teil von **► Tabelle 2** sind die Ergebnisse eingetragen. Es ergibt sich für Methode 2 ein geringer Unterschied in den Gesamtreichweiten: Soll = 15,05 und Ist = 14,93.

Abb. 1: Einfluss von r auf die Zielfunktion

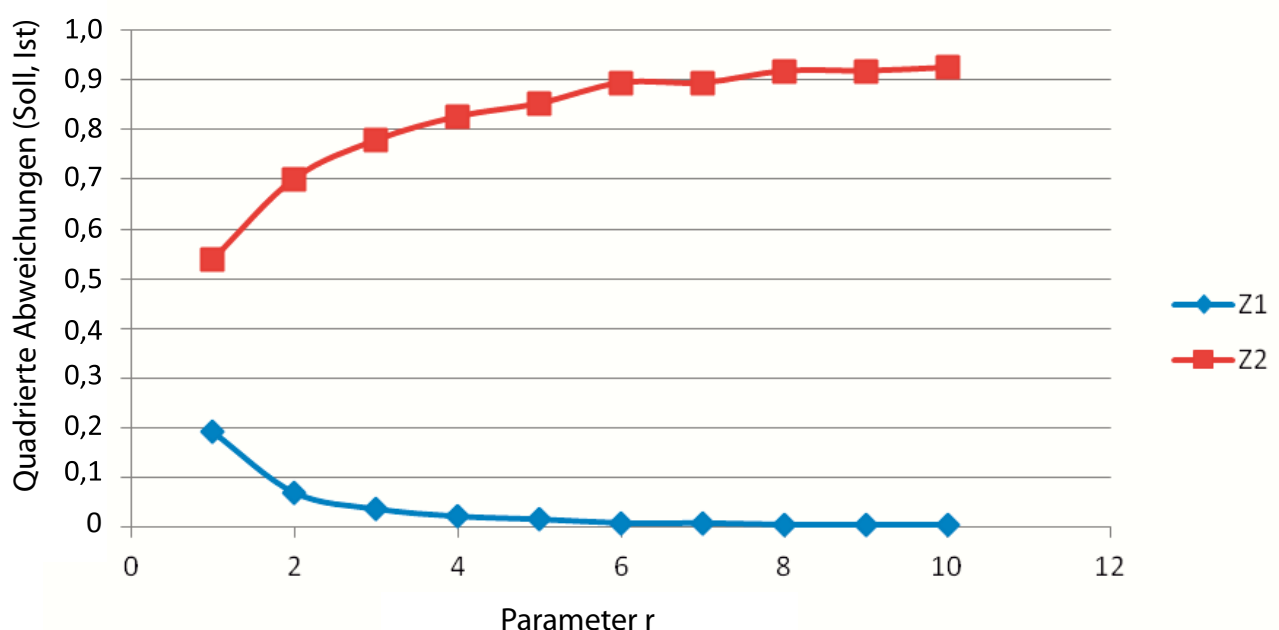
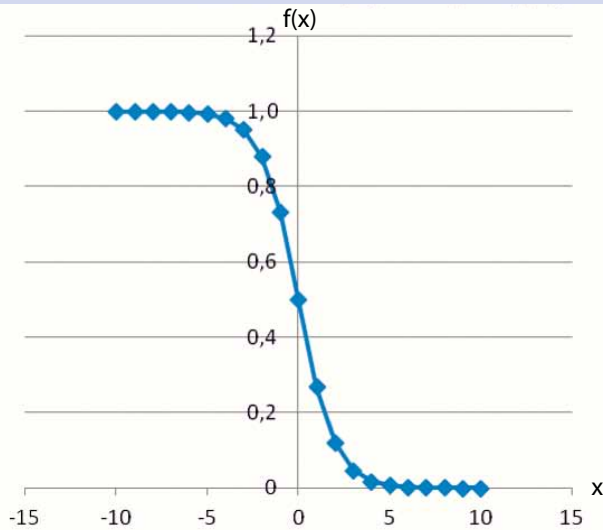


Abb. 2: Transformation $f(x)=a+(b-a)/(1+e^x)$ 

Dafür liegt die Lösung p_2 nun deutlich näher an den Originalwerten p_0 : 0,76 gegenüber 0,87 (Methode 1). Wieder haben die Personen 9, 12 und 18 unzulässige Wahrscheinlichkeitswerte über 1.

3.3 Die Methode mit Kontrollparameter und iterativer Anpassung der p-Werte

Bei dieser Methode, die nach dem Modell von Methode 2 arbeitet, liegen alle berechneten p-Werte im Bereich zwischen 0 und 1. Dies wird durch eine Transformation f erreicht (vgl. ▶ Abbildung 2), die beliebige reelle Zahlenwerte in einen Bereich von a bis b transformiert. Wählt man $a = 0$ und $b = 1$, so ergeben sich zwangsläufig Zahlen zwischen 0 und 1, womit das Problem der Methoden 1 und 2 nicht mehr auftritt.

Hier wurde mit $r = 4$ gerechnet (genauso wie bei Methode 2). Im unteren Teil von ▶ Tabelle 2 erkennt man, dass der Anpassungsprozess recht gut gelingt und obendrein die Lösung p_3 nah an p_0 liegt (geringfügig besser als bei Methode 1: 0,83 vs. 0,87). Besonders wichtig ist nun aber die Tatsache, dass alle Befragten diesmal zulässige Nutzungswahrscheinlichkeiten haben: Siehe die Werte p_3 in ▶ Tabelle 1.

Methode 3 verwendet ebenfalls das Modell der Zielfunktion $Z = r Z_1 + Z_2$, wobei nun allerdings die Funktionen Z_1 und Z_2 wegen der Transformation f nicht mehr quadratisch (wie bei Methode 2) sind. Deshalb kann das Minimum von Z nur noch iterativ ermittelt werden (Gradientenverfahren).

Das Iterationsverfahren lässt sich so skizzieren. Man beginnt mit einem im Prinzip willkürlichen Satz von n Zahlenwerten $x = (x_1, \dots, x_n)$. Anschließend werden diese

Werte durch die Rechenprozedur (im Anhang Formel 9) verändert. Die dadurch veränderten Werte in x werden abermals in dieselbe Prozedur eingesetzt und damit erneut verändert usw. Dieser Prozess wird solange fortgesetzt bis keine Veränderung bei x mehr feststellbar ist. Mit diesem x lautet die Lösung dann $p_3 = f(x)$. Ausführliche Details findet der interessierte Leser im Anhang (Formeln 8 und 9). Der Einfluss des Kontrollparameters r wird durch ▶ Abbildung 1 dokumentiert.

In ▶ Abbildung 1 sind die (quadrierten) Abweichungen zwischen den Soll- und den Ist-Werten der Gesamtreichweiten Z_1 und den Summen der (quadrierten) Abweichungen Z_2 der angepassten p-Werte (in p_3) und deren Originalwerten (in p_0) in Abhängigkeit von r eingetragen.

Man erkennt, dass mit zunehmenden Werten des Kontrollparameters, der Unterschied zwischen den Reichweiten verschwindet ($Z_1 \rightarrow 0$). Gleichzeitig aber entfernen sich die angepassten p-Werte von den Originalwerten p_0 (Anstieg von Z_2 bei wachsendem Parameter). Der gewählte Kontrollwert $r = 4$ in unseren Berechnungen scheint hier angemessen zu sein: während Z_1 für $r > 4$ kaum noch abfällt, steigt Z_2 weiter an. ▶ Abbildung 2 zeigt den Verlauf der Funktion f , die dafür sorgt, dass die p-Werte im Bereich 0 bis 1 liegen. Dies erfolgt durch die Wahl der Parameter $a = 0$ und $b = 1$.

Zur Berücksichtigung von Nutzergruppen bei der Anpassung (siehe Einführung) lässt sich Methode 3 nacheinander mit den jeweiligen Parametern a und b anwenden. Damit können zusätzlich die Reichweiten innerhalb der Nutzergruppen übertragen werden.

4. Fazit

Für Werbungtreibende und Mediaagenturen sind allgemein anerkannte Richtgrößen für Medien für Planungszwecke unverzichtbar. Als solche haben sich im Laufe der Jahre in Deutschland insbesondere die Reichweiten der Mediaanalyse der Arbeitsgemeinschaft Media-Analyse e.V. (ag.ma) als allgemein anerkannte Standards herausgestellt.

Bei kleineren Befragungen muss deshalb in der Regel eine nachträgliche Anpassung der Nutzungswahrscheinlichkeiten zur Sicherstellung dieser Normen erfolgen.

Im Aufsatz wurden hierzu drei Methoden vorgestellt, von denen der iterativen Methode 3 nach dem Gradientenverfahren der Vorzug gegeben wird. Für die Praxis ist besonders wichtig, dass sich damit die Reichweiten in beliebig vielen Zielgruppen optimal justieren lassen.

Mathematischer Anhang

Methode Nr. 1:

Vorbemerkungen: Alle fettgekennzeichneten Buchstaben sind Matrizen (Großbuchstaben) bzw. Spaltenvektoren (Kleinbuchstaben). Der Exponent T bedeutet das Vertauschen von Zeilen und Spalten einer Matrix bzw. den Übergang des Spaltenvektors \mathbf{x} zum Zeilenvektor \mathbf{x}^T (Transposition). Die Anzahl der Befragten sei n , die der Zielgruppen m . \mathbf{E}_n bzw. \mathbf{E}_m seien die Einheitsmatrizen der Ordnung n bzw. m .

Die Beweise für nachfolgende Formeln findet man in einem früheren Aufsatz (Kopp 2006).

Notationen: $\mathbf{p0}$ sei der Vektor der ursprünglichen (noch nicht angepassten) p -Werte, \mathbf{J} die Indikatormatrix der beteiligten Merkmale (im Beispiel entspricht diese der Matrix aus Nullen und Einsen von ▶ Tabelle 1, wobei die Zeilen mit den Spalten vertauscht sind, also eine Matrix der Ordnung 6×20). \mathbf{D} sei die Diagonalmatrix, die die Gewichte der Befragten enthält:

$$\mathbf{D} = \text{diag}(g_1, \dots, g_n). \mathbf{S} \text{ sei } \mathbf{JD}.$$

Hinweis: Sind alle Gewichte gleich 1, wie in unserem Beispiel, so gilt $\mathbf{S} = \mathbf{J}$.

Mit dem Vektor \mathbf{w} der Soll-Reichweiten in den m Zielgruppen erhält man die Optimierungsaufgabe:

$$\text{Min } (\mathbf{p} - \mathbf{p0})^T (\mathbf{p} - \mathbf{p0}) \text{ unter der Nebenbedingung } \mathbf{Sp} = \mathbf{w} \quad (1)$$

Gesucht wird hier der p -Wert-Vektor $\mathbf{p} = (p_1, \dots, p_n)^T$, der möglichst nah am Vektor $\mathbf{p0}$ der ursprünglichen Nutzungswahrscheinlichkeiten liegt und als Restriktionen die Gleichheit der Zielgruppenreichweiten \mathbf{Sp} mit den Sollvorgaben \mathbf{w} erfüllt. Der Ausdruck $(\mathbf{p} - \mathbf{p0})^T (\mathbf{p} - \mathbf{p0})$ ist der quadrierte Euklidische Abstand der Vektoren \mathbf{p} und $\mathbf{p0}$.

Die Lösung von (1) ist

$$\mathbf{p} = \mathbf{p0} + \mathbf{S}^T (\mathbf{SS}^T)^{-1} [\mathbf{w} - \mathbf{Sp0}] \quad (2)$$

Sonderfall: Obige Formeln und die nachfolgenden vereinfachen sich, wenn es nur um die Gesamtreichweite bei der Anpassung geht, nicht aber um Reichweiten einzelner Zielgruppen. In diesem Fall besteht \mathbf{J} aus einer einzigen Zeile aus lauter Einsen. Dementsprechend ergibt sich dann aus (2) für den p -Wert des Befragten j :

$$p_j = p0_j + k g_j [w - w0] \quad (3)$$

Hier ist $w0 = \sum_{i=1}^n g_i p0_i$ die Gesamtreichweite aus den Originalwerten $\mathbf{p0}$ und w der Sollwert der Gesamtreichweite. Für den Faktor k gilt:

$$k = 1 / \sum_{i=1}^n g_i^2$$

Insbesondere würde man für den p -Wert des 1. Befragten erhalten:

$$p_1 = p0_1 + k g_1 [w - w0] = 0,65 + 1/20 [15,05 - 13,33] = 0,74.$$

Methode Nr. 2:

Gelöst wird hier die unbeschränkte Minimierungsaufgabe $\text{Min } \{r (\mathbf{Sp} - \mathbf{w})^T (\mathbf{Sp} - \mathbf{w}) + (\mathbf{p} - \mathbf{p0})^T (\mathbf{p} - \mathbf{p0})\}$ (4)

In (4) ist r ein positiver Kontrollparameter. Mit wachsendem r konvergiert \mathbf{Sp} gegen \mathbf{w} – bei sinkenden Werten von r nähert sich \mathbf{p} an $\mathbf{p0}$.

Die Lösung von (4) ist

$$\mathbf{p} = \mathbf{p0} + \mathbf{S}^T (\mathbf{E}_m / r + \mathbf{SS}^T)^{-1} [\mathbf{w} - \mathbf{Sp0}] \quad (5)$$

Diese Formel vereinfacht sich im Sonderfall (Anpassung lediglich der Gesamtreichweite) zu

$$p_j = p0_j + k g_j [w - w0] \quad (6)$$

wobei nun $k = 1 / (r + \sum_{i=1}^n g_i^2)$ gilt. Für Person 1 ergäbe das:

$$p_1 = p0_1 + k g_1 [w - w0] = 0,65 + [1/(0,25+20)] [15,05 - 13,33] = 0,73.$$

Methode Nr. 3:

Die Funktion $f(x) = a + (b - a) / (1 + e^x)$ transformiert den Wert x in den Bereich (a, b) .

Für Vektoren der Art $\mathbf{x} = (x_1, \dots, x_n)^T$ gilt die Vereinbarung $f(\mathbf{x}) = (f(x_1), \dots, f(x_n))^T$.

Die Aufgabe besteht nun in der Minimierung der Zielfunktion

$$Z = r (\mathbf{Sf}(\mathbf{x}) - \mathbf{w})^T (\mathbf{Sf}(\mathbf{x}) - \mathbf{w}) + (\mathbf{f}(\mathbf{x}) - \mathbf{p0})^T (\mathbf{f}(\mathbf{x}) - \mathbf{p0}). \quad (7)$$

Die Funktion Z ist von der Art $Z = r Z1 + Z2$, wobei $Z1$ für die Summe der quadrierten Abweichungen zwischen den Soll-Reichweiten \mathbf{w} und den Ist-Reichweiten $\mathbf{Sf}(\mathbf{x})$ steht. $Z2$ misst den quadrierten Euklidischen Abstand der Vektoren $\mathbf{f}(\mathbf{x})$ und $\mathbf{p0}$.

Der Gradient des Ausdrucks (7) lautet

$$\text{Grad } Z = 2 \text{ diag} (f'(\mathbf{x})) \{ [r \mathbf{S}^T \mathbf{S} + \mathbf{E}_n] f(\mathbf{x}) - r \mathbf{S}^T \mathbf{w} - \mathbf{p} \}. \quad (8)$$

Hierbei ist $f'(\mathbf{x}) = (a - b) e^x / (1 + e^x)^2$ die Ableitung von $f(\mathbf{x})$.

Damit erfolgt die Lösung von (7) per Gradienten-Verfahren 1. Ordnung gemäß

$$\mathbf{x} = \mathbf{x} - s \text{ Grad } Z \quad (9)$$

In (9) ist s eine positive Schrittweite. Nach Abschluss der Iterationen sind die gesuchten Nutzungswahrscheinlichkeiten durch $\mathbf{p} = f(\mathbf{x})$ gegeben.

Zur Technik der Gradientenverfahren siehe Meister (2005) und Zurmühl (1965).

Literatur

Büttner, J. (1979): Ein Verfahren zur Anpassung der Lesewahrscheinlichkeiten zweier Mediaanalysen, in: Zeitschrift für Operations Research, 23 (6), B171 – B179.

Kopp, B. (2006): Eine Methode zur Strukturanpassung von Umfragedaten, in: ZUMA-Nachrichten, 59, 89 – 110 (<http://nbn-resolving.de/urn:nbn:de:0168-ssoar-69267>).

Meister, R. (2005): Numerik linearer Gleichungssysteme, 2. Aufl., Vieweg, Wiesbaden.

Müller-Merbach, H. (1973): Operations Research, Vahlen, München.

Zurmühl, R. (1965): Praktische Mathematik für Ingenieure und Physiker, 5. Aufl., Springer, Berlin, Heidelberg, New York.

Impressum

57. Jahrgang, 228. Folge

Herausgeber:

Deutsche Werbewissenschaftliche Gesellschaft e.V.
Kleine Johannisstraße 6, 20457 Hamburg, Deutschland
Tel.: +49 (0) 40 3060459-16, Fax: +49 (0) 40 3060459-29
info@dwg-online.net, www.dwg-online.net

Österreichische Werbewissenschaftliche Gesellschaft
Augasse 2-6, 1090 Wien, Österreich
Tel.: +43/1/313 36-4617, Fax: +43/1/317 66 99
wwg@wu.ac.at, www.wgonline.at

Verlag:

New Business Verlag GmbH & Co. KG
Nebendahlstraße 16, 22041 Hamburg, Deutschland
Tel.: +49/40/60 90 09-0, Fax: +49/40/60 90 09-66
transfer@new-business.de

Verleger: Peter Strahlendorf

Kfm. Geschäftsführung: Antje-Betina Weidlich-Strahlendorf

Produktmanagement: Anja Kruse-Anyaegbu

Vertrieb: Angelika Schmidt

Gestaltung: Klaus Esche

Entwurf: Nina Leinemann

Druck: Lehmann Offsetdruck GmbH, Norderstedt

Redaktion dieser Folge:

DWG

Für die DWG:

Chefredakteure: Prof. Dr. Martin Eisend, Europa-Universität Viadrina
Prof. Dr. Bernhard Heidel, Hochschule RheinMain, Wiesbaden Business School
Prof. Dr. Tobias Langner, Bergische Universität Wuppertal
alle V.i.S.d.P.

Für die WWG:

Chefredakteur: Prof. Dr. Günter Schweiger V.i.S.d.P.
Stv. Chefredakteur: Prof. Dr. Wolfgang Mayerhofer
Redaktion: Mag. Thomas Biruhs
alle Wirtschaftsuniversität Wien

Anzeigenverwaltung:

Österreich: Österreichische Werbewissenschaftliche Gesellschaft
Verantwortlich für Anzeigen: Mag. Thomas Biruhs, Mag. Beatrice Buzarna

Deutschland: New Business Verlag

Verantwortlich für Anzeigen: Milosz Lipski, lipski@new-business.de

Anzeigendisposition: Silke Reyher-Timmann

Herausgeberbeirat:

Prof. Dr. Carsten Baumgarth, Hochschule für Wirtschaft und Recht Berlin
Prof. Dr. Dirk-Mario Boltz, Hochschule für Wirtschaft und Recht Berlin
Prof. Dr. Manfred Bruhn, Universität Basel
Prof. Dr. H. Dieter Dahlhoff, Universität Kassel
Prof. Dr. Franz-Rudolf Esch, EBS Business School, Oestrich-Winkel
Prof. Dr. Thomas Foscht, Universität Graz
Dr. Gereon Friederes, marketmind Markt- und Markenforschung, Wien
Prof. Dr. Michael Grabner, Michael Grabner Media, Wien
Prof. Dr. Andrea Gröppel-Klein, Universität des Saarlandes, Saarbrücken
Prof. Dr. Arnold Hermanns, Universität der Bundeswehr, Neubiberg
Karen Heumann, Jung von Matt AG, Hamburg
Siegfried Högl, GfK Marktforschung, Nürnberg
Dr. Helene Karmasin, Karmasin Motivforschung, Wien
Prof. Dr. Peter Kenning, Zeppelin Universität, Friedrichshafen
Prof. Dr. Erich Kirchler, Universität Wien
Prof. Dr. Guido Kucsco, Rechtsanwalt in Wien
Mag. Willy Lehmann, Willy Lehmann Markenagentur, St. Florian
Prof. Dr. Wilfried Leven, Agentur Leven, Köln
Prof. Dr. Josef Mazanec, Wirtschaftsuniversität Wien
Dr. Walter Neuhauser, Neuhauser Harnach GbR, München
Prof. Dr. Bruce I. Newman, De Paul University, Chicago
Dr. Oliver Nickel, Icon Added Value, Nürnberg
Prof. Dr. Thomas Otter, Goethe Universität, Frankfurt
Dr. Thomas Petersen, Institut für Demoskopie Allensbach, Allensbach
Prof. Dr. Bodo Schlegelmilch, Wirtschaftsuniversität Wien
Prof. Dr. Peter Schnedlitz, Wirtschaftsuniversität Wien
Prof. Dr. Günter Schweiger, Wirtschaftsuniversität Wien
Prof. Dr. Gerhard Speckbacher, Wirtschaftsuniversität Wien
Prof. PD Dr. Andreas Strebing, York University, Toronto
Prof. Dr. Peter Szyszka, Universität Wien
Prof. Dr. Volker Trommsdorff, Technische Universität Berlin
Prof. Dr. Gerhard A. Wührer, Johannes Kepler Universität Linz

www.transfer-zeitschrift.net

Abonnentenpreis (4 Ausgaben):

78,- Euro pro Jahr; Hochschulangehörige: 65,- Euro pro Jahr; Bibliotheken: 50,- Euro pro Jahr. Für Mitglieder der angeschlossenen Gesellschaften ist der Abonnentenpreis im Mitgliedsbeitrag enthalten.

Postvertriebszeichen: G42362 (Österreich), C43262 (Deutschland)

ISSN 1436-798-X, Copyright © 2011 DWG/WWG

Nachdruck von Beiträgen mit Quellenhinweis und gegen ein Belegexemplar nur mit ausdrücklicher Genehmigung von DWG/WWG.